Available online at www.sciencedirect.com

ScienceDirect

Procedia Technology 24 (2016) 15 – 22

Procedia Technology

International Conference on Emerging Trends in Engineering, Science and Technology (ICETEST - 2015)

Regionalization of Rainfall in Kerala State

Shahana Shirin A H [a], Reeba Thomas[b]*

[a] *Student, Dept. of Civi Engineering, Govt. Engineering College, Trichur-680009, Kerala, India*
[b*] *Associate professor, Dept. of Civi Engineering, Govt. Engineering College, Trichur-680009, Kerala, India*

Abstract

The main objective of this study is to classify the entire State of Kerala into optimum number of homogeneous zones based on monthly rainfall. The same is done by Ward's Clustering method by taking the monthly rainfall data for five consecutive years from 2008-2012 of 63 raingauge stations in Kerala state. Clustering is the process of dividing the area under consideration to a limited number of climatologically homogeneous zones, based on any hydrologic parameter. The purpose of cluster analysis is to place objects into groups suggested by the data, so that objects in a given cluster tend to be similar to each other and objects in different clusters tend to be dissimilar. It is found that the region can be grouped into seven clusters. There exists large variability in average annual rainfall of different clusters. There are cluster groups where the annual rainfall is as high as 1.5 times and as less as half of average annual rainfall of Kerala as a whole.

Peer-review under responsibility of the organizing committee of ICETEST – 2015

*Keywords:*Hierarchical clustering; Ward's; Regionalization; Kerala.

1. Introduction

Kerala experiences rainfall for around 120 to140 rainy days per year. Climate of Kerala state is influenced by the seasonal heavy rains of the southwest summer monsoon and northeast winter monsoon. Around 65% of the rainfall occurs during southwest monsoon and the rest during north east monsoon. The hills and mountains of the Western

* Corresponding author. Tel.: +919447046464.
*E-mail address:*reeba@gectcr.ac.in

Peer-review under responsibility of the organizing committee of ICETEST – 2015
doi:10.1016/j.protcy.2016.05.004

Ghats located on the eastern boundary of the State provide orographic lifting for the southwest monsoon winds resulting in heavy precipitation over the western slopes and good rain over mid and low lands. Northeast monsoon also contribute to the annual rainfall, especially in the southern parts of the State.

Though, the state of Kerala is blessed with 44 rivers, the inhabitants of major part of it experience water scarcity in summer season. During this period, most of these rivers remain as still water pools or marsh lands. Due to the population increase, rapid industrialization and unhealthy anthropologic interventions, there is a sharp decline in the level of water table also. Thus there exists a pressing need for a sustainable methodology to develop watershed so as to manage the optimum use of available water resources round the year. For sustainable accomplishment of watershed management, climatic parameters of the entire basin are to be known which consist of both gauged and ungauged sub- basins. The parameters at the ungauged sub-basins can be made available via regionalization. And hence the technique of regionalization gains significance by which similar watershed basins can be identified.

[5] has observed that in order to get a good regionalization, it is important to answer two questions: i) how many clusters are required according to the type of data or objective and ii) how good is the clustering itself. To answer these questions, hierarchical and non-hierarchical algorithms were applied to six experiments based on the data sets for precipitation and temperature available from traditional weather stations.

[4] used Principal Component Analysis to divide Spain in to a number of climatically homogeneous zones. Here Ward's method is used for clustering. Here, optimum number of clusters is determined by plotting distance between merged clusters as a function of the stage analysis.

[1] has conducted clustering analysis of Australia using Principal Component Analysis and k. means cluster analysis. Results have shown that the regions identified by both analyses are largely same.

In [3], the entire region of Bangladesh is grouped into certain homogeneous regions based on temperature and rainfall by using K mean clustering method.

In this paper an attempt has been made to group the entire region of Kerala into an optimum number of clusters using Ward's method of clustering. The purpose of Clustering Analysis is to place objects into groups suggested by the data, not defined previously, so that objects which are similar to each other in some sense are grouped into one cluster, and dissimilar objects in different clusters.

Nomenclature

P	number of clusters
N	number of elements (gauging stations) in each cluster
M	number of variables (months)
X_{kij}	j^{th} month at the i^{th} gauging station in the k^{th} cluster
$\bar{X}_{kj}$	average value of the i^{th} gauging station at the j^{th} month in the k^{th} cluster

2. Study Area and Data Description

For this study, the entire state of Kerala, which occupies the southernmost part of India, is selected. Kerala is a small coastal State, situated in the southwest peninsular part of India. It lies between a latitude of 8^0 18'N and 12^0 48'N and a longitude of 74^0 52'E and 77^0 22'E. The total area of the state accounts to around 38863 sq.Km, with a coastal belt of around 600km. Kerala state is endowed with rich water resources like 44 rivers and two important

monsoons. However, the high density of population has exerted pressure on water resources for meeting the various needs of the people.

For the study, the main parameter required is the monthly rainfall of all the gauging stations based on which the classification is done. The daily rainfall data are collected for a period of 5 years (2008-2012). 5 year rainfall data for 47 stations are obtained from Hydrology Sub Division, Jalavijnan Bhavan,Thiruvananthapuram. The data of remaining 16 stations could not be received as they lie in watersheds of interstate water disputes. Hence for these stations, Global Weather Data are used. They are downloaded from the site named, Global Weather Data for SWAT. Monthly rainfall is averaged over the period of availability of data to obtain the average monthly rainfall and it is used for Ward's clustering method.

3. Methodology

3.1. Hierarchical Clustering (Ward's Method)

Average monthly rainfall data for 5 years of 63 stations are used for Ward's method. 63 × 12 matrix is given as input Data matrix (X) in MATLAB. Assuming that, there are 'n' elements to cluster. The process of clustering initially begins with 'n' clusters each with one element. Then numbers of classes are reduced by merging the similar pairs of clusters. The pairwise distance between pairs of objects that is, Euclidean distance between pairs of object is computed by using 'pdist' function in MATLAB. It is a row vector of n (n-1)/2 sizes. Then it is converted into square matrix of size 63 × 63, to save space and computational time. Then created agglomerative hierarchial cluster tree from the distance, which is known as dendrogram. Dendogram is a bi-dimensional figure that represents the different stages and the distance at which the observations are clustered. Number of clusters and number of elements in each cluster is determined by using dendrogram.

3.2. Determination of Optimum Number of Clusters

Dendrogram obtained from Ward's algorithm is used for the determination of optimum number of clusters. When similar clusters are being merged early in the process, these distances are small, and they increase relatively little by little as the clustering proceeds. The process can be stopped just before these distances become large. The optimum number of clusters found by the dendrogram based on the vertical heights has to be checked for its accuracy by some other reliable methods. For which a statistical factor called as Root Mean Square Standard Deviation (RMSSTD) is used. RMSSTD is the mean square root of the variance of each variable of the new cluster formed by merging two clusters and summed over all variables. It's a constant value for each number of clusters to be formed. RMSSTD is represented mathematically as given in "(1)".

$$\text{RMSSTD} = \sqrt{\left(\frac{1}{M}\sum_{k=1}^{P}\sum_{i=1}^{N}\sum_{j=1}^{M}(x_{kij} - \bar{x}_{kj})^2\right)} \quad (1)$$

4. Analysis and Results

For this study monthly rainfall of 63 rain gauge stations are used. Locations of rain gauge stations are as shown in Fig.1. Daily rainfall data of 63 stations for the selected 5 year period (2008-2012) are obtained from Hydrology Sub division, Jalavijnan Bhavan, Thiruvananthapuram. Monthly rainfall is calculated from these daily rainfall data using MS Excel. Average monthly rainfall calculated for the 63 stations is used as input. To perform Ward's hierarchical clustering, a data matrix of size 63×12 is fed into MATLAB. The output for Ward's algorithms is visually interpreted using a Dendrogram shown in Fig.2. Optimum number of clusters is decided by noting the RMSSTD value and by selecting the one with the minimum RMSSTD value. RMSSTD value at different stage is given in Table 1.

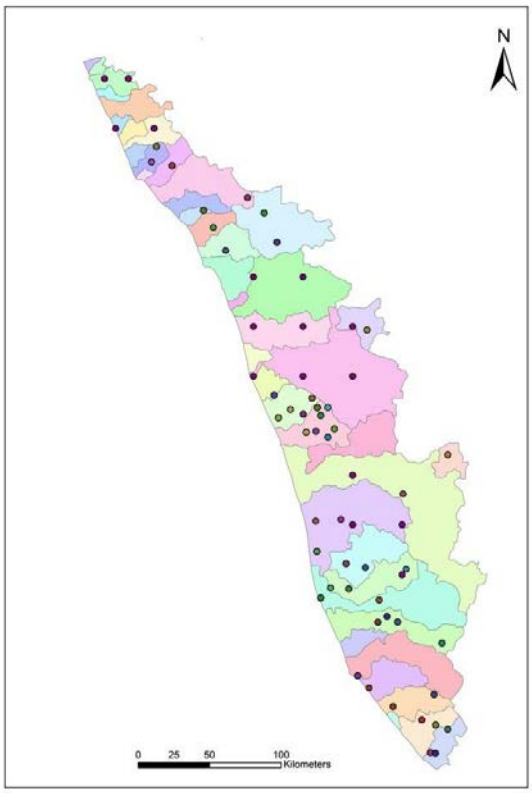

Fig.1. Location of Raingauge Stations in the Watersheds of Kerala

Table 1. RMSSTD value for different number of clusters

Number of clusters	RMSSTD
2	3030
3	2610
4	2369
5	2360
6	2290
7	2200
8	2880

On observation of the dendrogram, the optimum number of clusters is obtained by cutting the dendrogram at some level of clustering. RMSSTD value is calculated at each stage and plotted. Graph drawn between number of clusters and RMSSTD is given in Fig.8. Initially RMSSTD value decreases with the increase in the number of clusters. Then its value abruptly increases when the number of cluster is eight. ie. RMSSTD value is minimum when the number of cluster is seven. So, optimum number of cluster is selected as seven. Number of stations lying in each cluster as per Ward's Algorithm is given in Table 2.

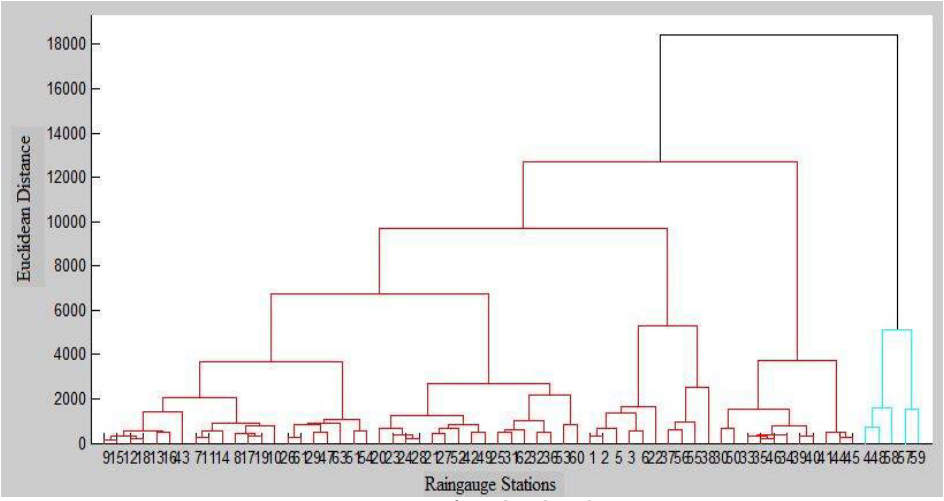

Fig.2. Output of Ward's Algorithm

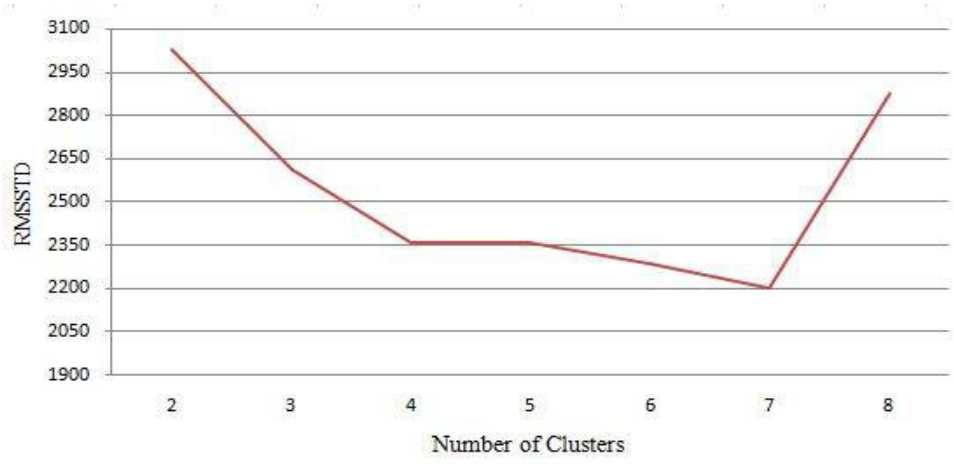

Fig.3. Graph of RMSSTD Vs Number of Clusters

Table 2. Output of Ward's Algorithm

Cluster Number	Number of stations in each cluster
1	14
2	7
3	16
4	6
5	4
6	11
7	5
Total	63

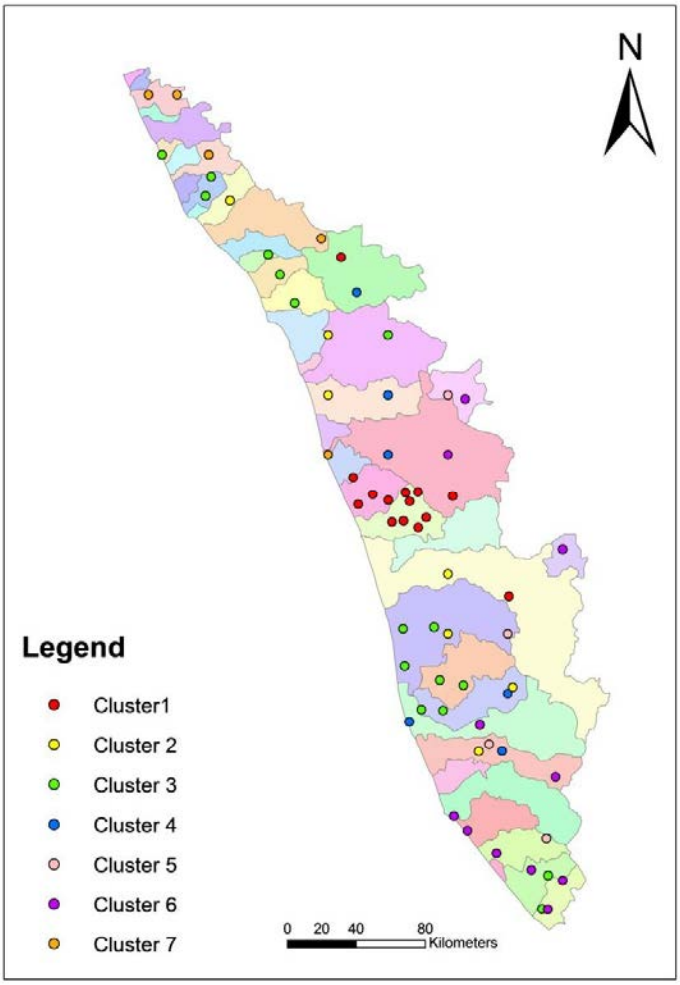

Fig.4 Cluster Map of Kerala based on Ward's Method

Homogeneous regions of Kerala under each cluster are identified by 'Thiessen Polygon Method' and the obtained map is shown in Fig.5.

Table 3. Average Annual Rainfall of Different Clusters

Cluster	Average Annual Rainfall(mm) (mm)
1	2638.46
2	3362.57
3	3767. 70
4	4358.48
5	2428.85
6	1501.36
7	4440.13

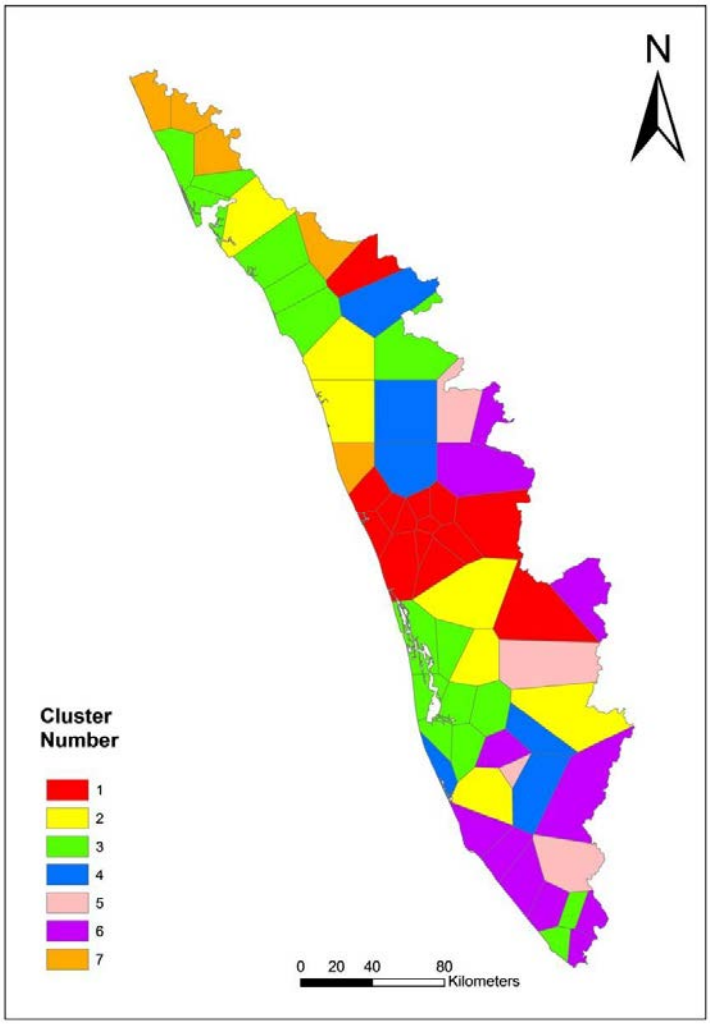

Fig.5. Homogeneous zones based on Ward's Method

Average annual rainfall of each cluster is calculated and tabulated in Table 3 and plotted in Fig. 6. Average annual rainfall is maximum in cluster 7 and 4, which is about 4400 mm. Minimum is in cluster 6 which is about 1500mm.

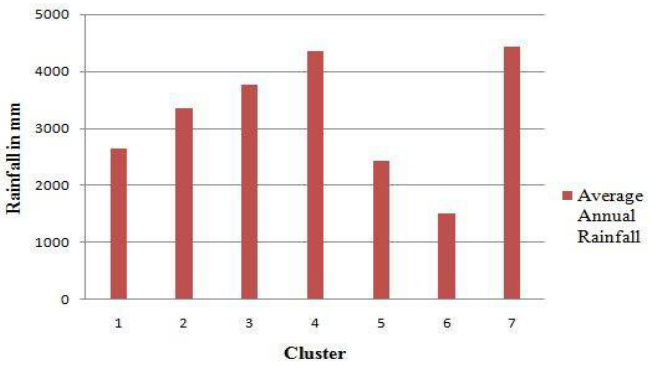

Fig.6. Average Annual Rainfall of different clusters

5. Conclusions

After conducting the study on Regionalization of Rainfall in Kerala state, several conclusions are drawn. Here Ward's clustering method is used on 5 year monthly rainfall for the period of 2008-2012. Optimum number of clusters obtained by Ward's algorithm is seven. Homogeneous regions of Kerala under each cluster are identified. Average annual rainfall is calculated for each cluster. Rainfall is maximum in cluster 7 and 4, which comes to about 4400 mm. ie. Mean rainfall of these groups is as high as 1.5 times of average annual rainfall of Kerala. Rainfall is minimum in cluster 6 and its magnitude is about 1500 mm and it is as less as half of average annual rainfall of Kerala. It indicates that, there is large variability in rainfall across Kerala.

Acknowledgements

The authors hereby acknowledge the financial support from TEQIP Phase II, Government Engineering College Trichur, Thrissur for seed money project "Variability Analysis and Regionalization of Rainfall in Kerala State".

References

[1] Baeriswyl P. A., Rebetez M. Regionalization of Precipitation in Switzerland by Means of Principal Component Analysis. Theoretical and Applied Climatology; 1996. p. 31-41.
[2] Fernandez Mills G. Principal Component Analysis of Precipitation and Rainfall Regionalization in Spain. Theoretical and Applied Climatology; 1994. p. 169-183.
[3] Md. Mizanur Rahman, Sarkar S., Reza Najafi M., Rai R. K. Regional Extreme Rainfall Mapping for Bangladesh Using L-Moment Technique. Journal of Hydrologic Engineering © ASCE; 2013. p. 603-615.
[4] Munoz-D´ıaz. D., Rodrigo F. S. Spatio-temporal patterns of seasonal rainfall in Spain (1912–2000) using cluster and Principal Component Analysis: comparison. Annales Geophysicae; 2004. p. 1435-1448.
[5] Ildiko Pelczer, Judith Ramos, Ramón Domínguez, Fernando González. Clustering of Regional Homogeneous Zones in a Watershed Using Clustering Algorithms. National Autonomous University of Mexico; 2007.
[6] Raziei T., Bordi I., Pereira L. S. A Precipitation based regionalization for Western Iran and regional drought variability. Hydrology and earth system sciences; 2008. p. 1309-1321.